

# Weekly Report

21<sup>st</sup>, Oct. - 26<sup>th</sup>, Nov. 2017

## Done:

1. Finish the courses and coding assignments on freecodecamp about Node.js and Express.js. Watch some online videos to enhance the understanding. Had a deep understanding of these concepts with practices.
2. Again, screened the provided links, finding possibly useful data. Knew the updating mechanism after consulting the department which provides the data. The useful data are listed as follows:

### **Air Quality**

1. AirNow Air Quality: historical data (With animation, every 20 minutes, params: O3, PM, max AQI )

From 2008 <http://www.sparetheair.com/histcalendar.cfm>

2. Air Quality Data (PST) query tool <https://www.arb.ca.gov/aqmis2/aqdselect.php>

From 1980, more types of params

### **Infrastructure**

1. Not accessible

### **Earthquake (may not be helpful)**

1. ShakeMap

### **Energy**

1. Electric Outages on a map (updated every 15 min)  
[https://m.pge.com/?WT.pgeac=Home\\_Outages#outages](https://m.pge.com/?WT.pgeac=Home_Outages#outages)

2. SMUD(Sacramento Municipal Utility District) real-time

### **Fire**

1. InciWeb
2. CalFire

### **HazMat**

1. oil spill, chemical release or maritime security. Records of all the calls by year, from 1990. <http://www.nrc.uscg.mil/>

### **California Schools**

1. a summary of public and private K-12 schools in California

### **Pipeline**

1. Gas transmission, hazardous liquid, and breakout tanks, accidents, incidents  
<https://pvnpm.phmsa.dot.gov/PublicViewer/> a query tool

### **Water, Flood related**

1. not helpful

### To Do:

1. Discuss about the currently available data of fire emergency.
- 2.

### Paper reading

KDD 17 *Meta-graph based recommendation fusion over heterogeneous information networks*  
heterogeneous information networks 简称 HIN, 是现在日常生活中非常常见的网络形式. 简单来说, 在推荐领域, 用户-item (产品店铺)这样的 bipartite graph 就是一种最简单的. 复杂的话还有把用户评论管检测, 签到地点, 用户关系, 产品分类等全部糅在图里. meta-graph 是相当于把复杂的图抽象化, 找出其 schema. 本文用各种抽象出来的不同 meta-graph 或者 meta-path(就是一条链), 利用数据建立从用户到 item 的相似性矩阵, 在对其做分解找出隐藏特征训练模型做推荐. 类似的异构图或者隐藏特征的可视化在最近 vis pvis eurovis 都有所出现, 虽然现在 NN 很厉害但估计速度上可能还是这个比较快. KDD 文章很多都需要做一些优化算法的设计和实现, 这一块我感觉我们很薄弱.

VIS 17 *Understanding a sequence of sequence: visual exploration of categorical states in lake sediment cores* 本文是对序列的序列做可视化, 虽然不复杂但好像是新的问题. 解决的核心问题是 domain specific 的, 描述的比较到位; 可视化方面比较简单, 只是依据用户需求做了比较适合的聚合以及有一些简单的推荐. 发在了 tvcg, 个人觉得是因为这个和领域问题结合紧密并且描述的比较好好.

VIS 17 *TreePOD: sensitivity-aware selection of pareto-optimal decision trees* 在之前的 VDS 有人做了随机森林的可视化然而给人觉得很一般, 本文不做森林, 只做一棵树, 覆盖了各个关键步骤. 其系统的使用者是领域专家但是不一定有统计或者机器学习方面的特别背景; 要从大量设定生成的候选中(叶子数量 树深度 最后树分裂点的值等等)一步步选出最适合的树. 树的可视化设计借鉴了之前的一篇文章, 但是其辅助视图我觉得做的很全面, 对于筛选真正有用的树所需的几乎可以说面面俱到. 他还将树用 treemap 表示出来, 不同颜色表示不同类型, 不同颜色点在一个 treemap 格子里用随机撒的方法可视化, 这个对于展现树不同层节点效果非常有用. 随机撒也是出于用户感知的角度, 若按序排布容易造成感知上的偏差. 本文就是单纯做一个针对机器学习方法的可视化, 确实做得很赞.

VIS 17 *Bring it to the Pitch: Combining video and movement data to enhance team* 文章最强的部分应该是在计算机视觉方面, 其最终结果就像是把一场真正的足球比赛视频做成了足球游戏的比赛一样, 其实很多东西在足球游戏里都有可视化. 并且其高超的技术能让一个可视化编码出现在二维的投影和三维视频中都出现的比较自然. 之前有棒球的可视化文章做过类似的工作但不是直接在视频上做的. 这样的可视化结果也是希望能研究战术\做一些预测, 可以说在兰吉工作上也可以加一些类似的东西, 工作量相比起来会小一点.

安排表

内容	DDL	Milestone
设计搭建框架	12.10	仍然在整合数据. 比较复杂
前后端学习	A.S.A.P	学习了 <code>nodejs</code> 和 <code>expressjs</code> , 完成一些习题并看了一些视频巩固